



# Why Data Migration Projects Fail

Joe Hudicka





# What is Data Migration?

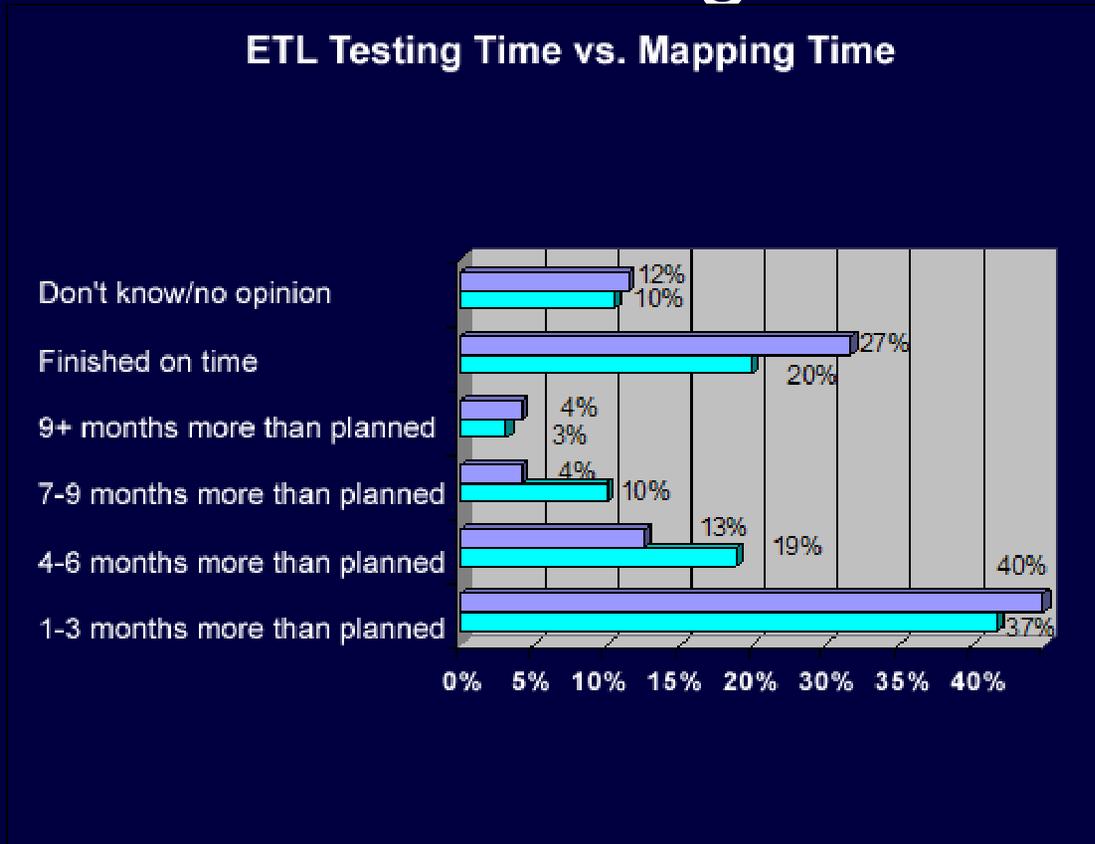
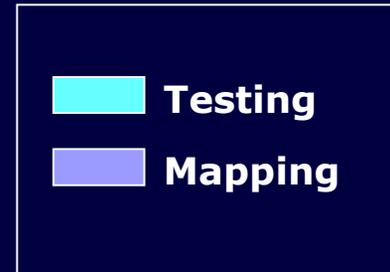
- The set of processes by which data is retrieved, transformed and transferred between two or more applications
  - Data warehousing/Business intelligence
  - Business process reengineering efforts
  - Inter/Intra-organizational interfaces





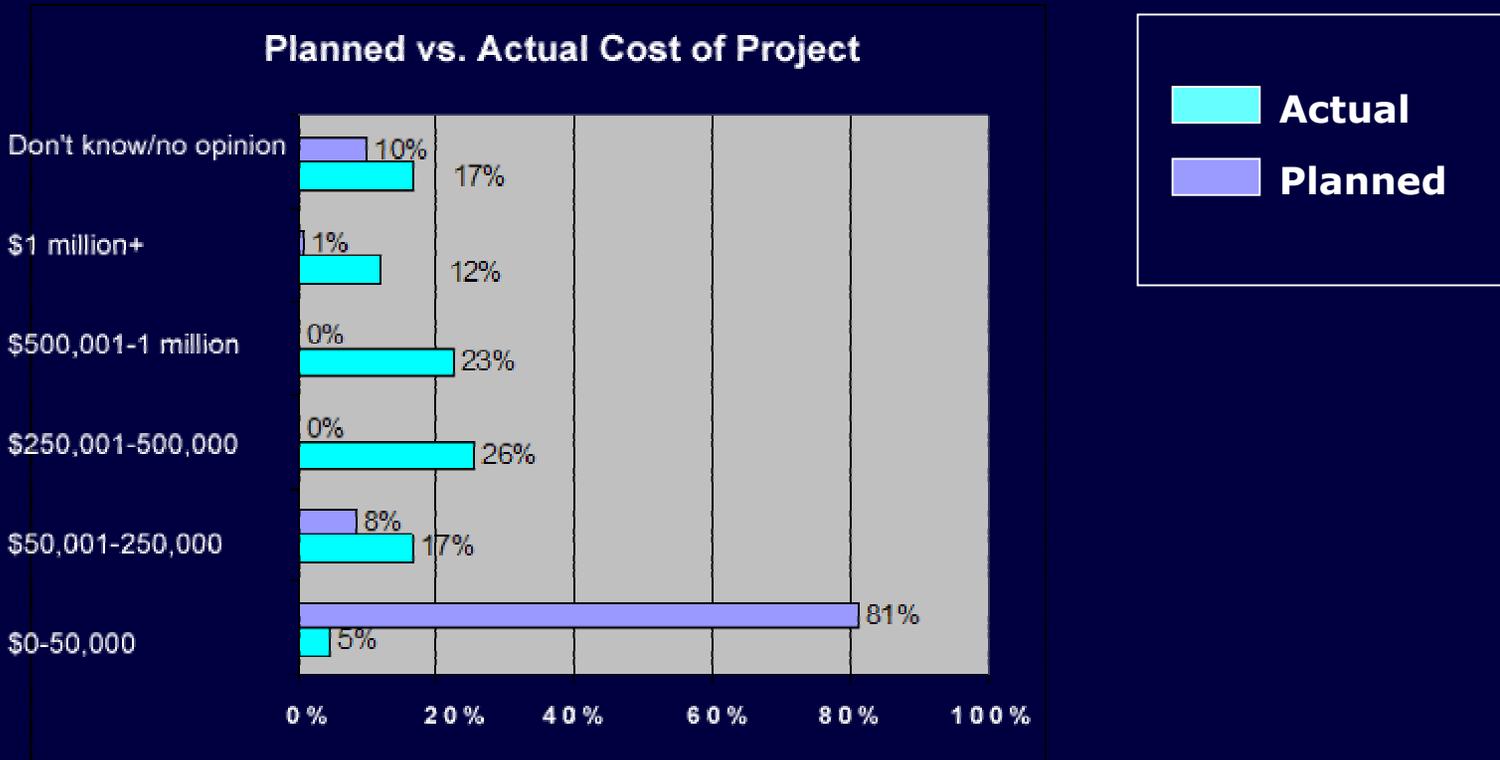
# Over 67% of Migrations Finished Late

## ETL Testing Time vs. Mapping Time





# Migrations Cost an Average of 14 Times the Planned Amount





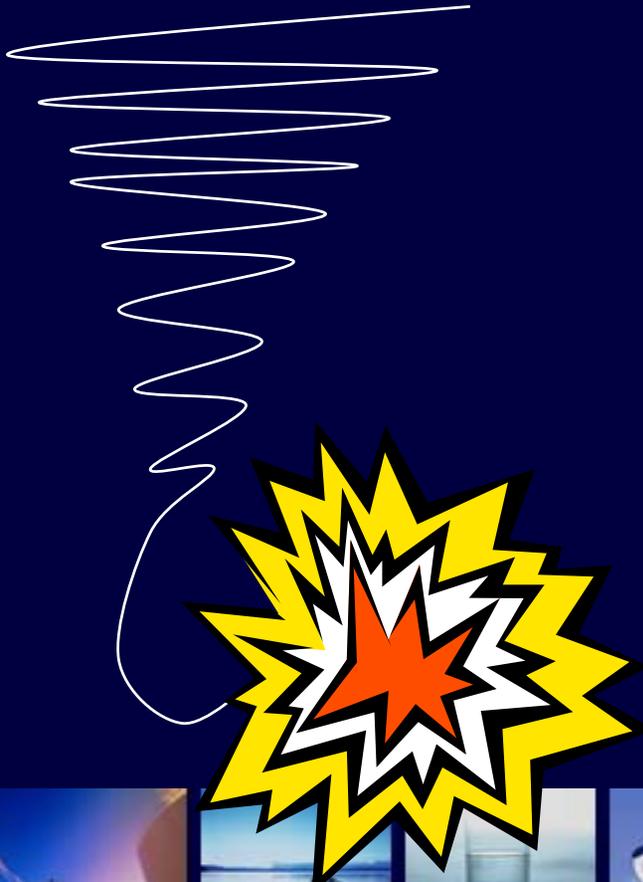
# Data Migration Obstacles

- Disparate data source designs/technologies
- Limited data validation enforced
- Little historic knowledge of data available
- Lack of clarity regarding records retention policies
- General assumption that all data is “Good Data”





# Spiral Methodology (Code, Load & Explode!)



- Ignores data quality
- Over budget
- Out of time
- Poor performance
- Trivial error management



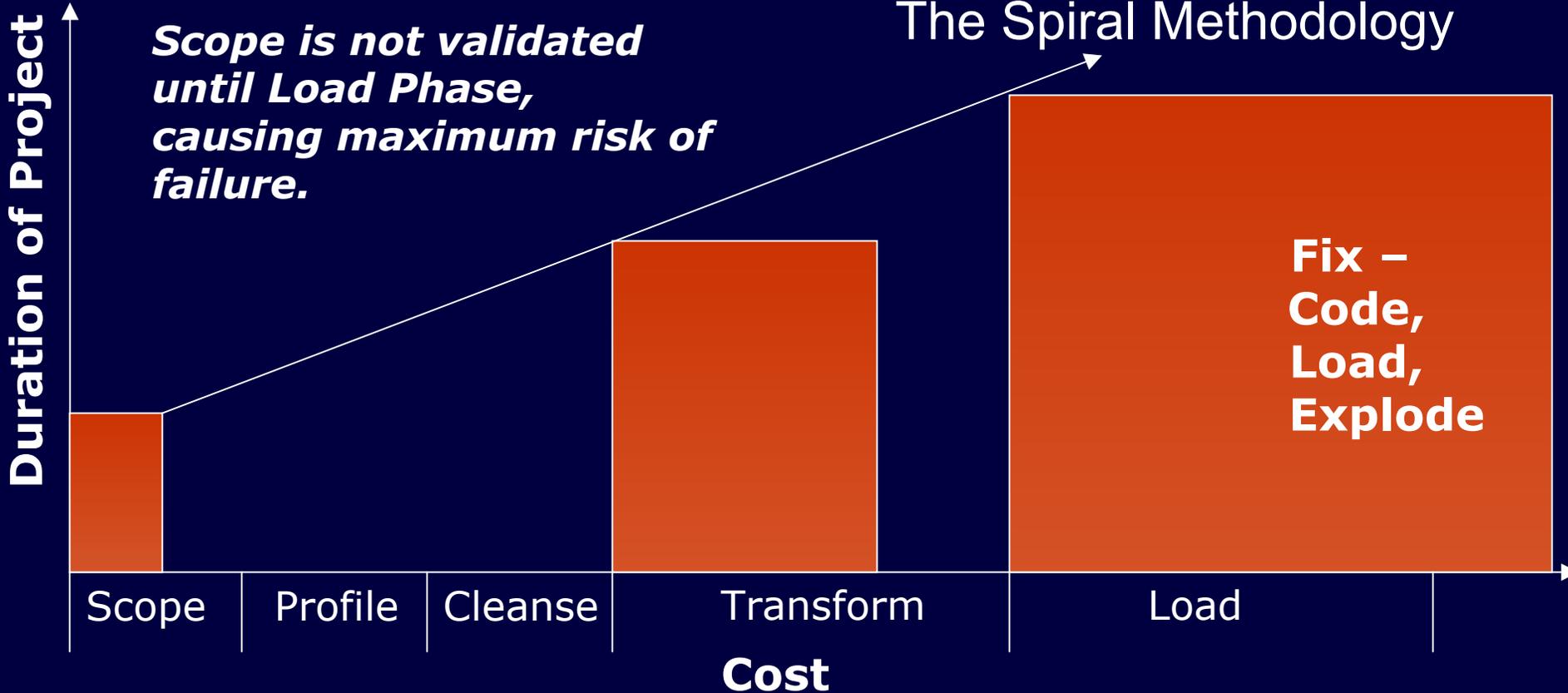


# The Traditional Approach

The Spiral Methodology

*Scope is not validated until Load Phase, causing maximum risk of failure.*

**Fix –  
Code,  
Load,  
Explode**





# Why Does the Code Explode?

Extract  
Rules

Transform  
Rules

Load  
Rules

Cleansing  
Rules

Cleansing  
Rules

~~ETL~~





# What's Missing?

- Understanding of the strengths — and weaknesses — of pre-existing data
- Ability to cleanse, standardize and validate data throughout the migration process





# The Data Profiling and Quality Challenge

## Time Consuming

One Organization needed 18 months to complete data quality analysis and cleansing

## No Documentation

Relationships across tables may be missing  
Metadata may be incorrect

## Difficult

Robert Smith Jr → Bob Smith → Mr. Smith, Robert  
Writing code is difficult and expensive



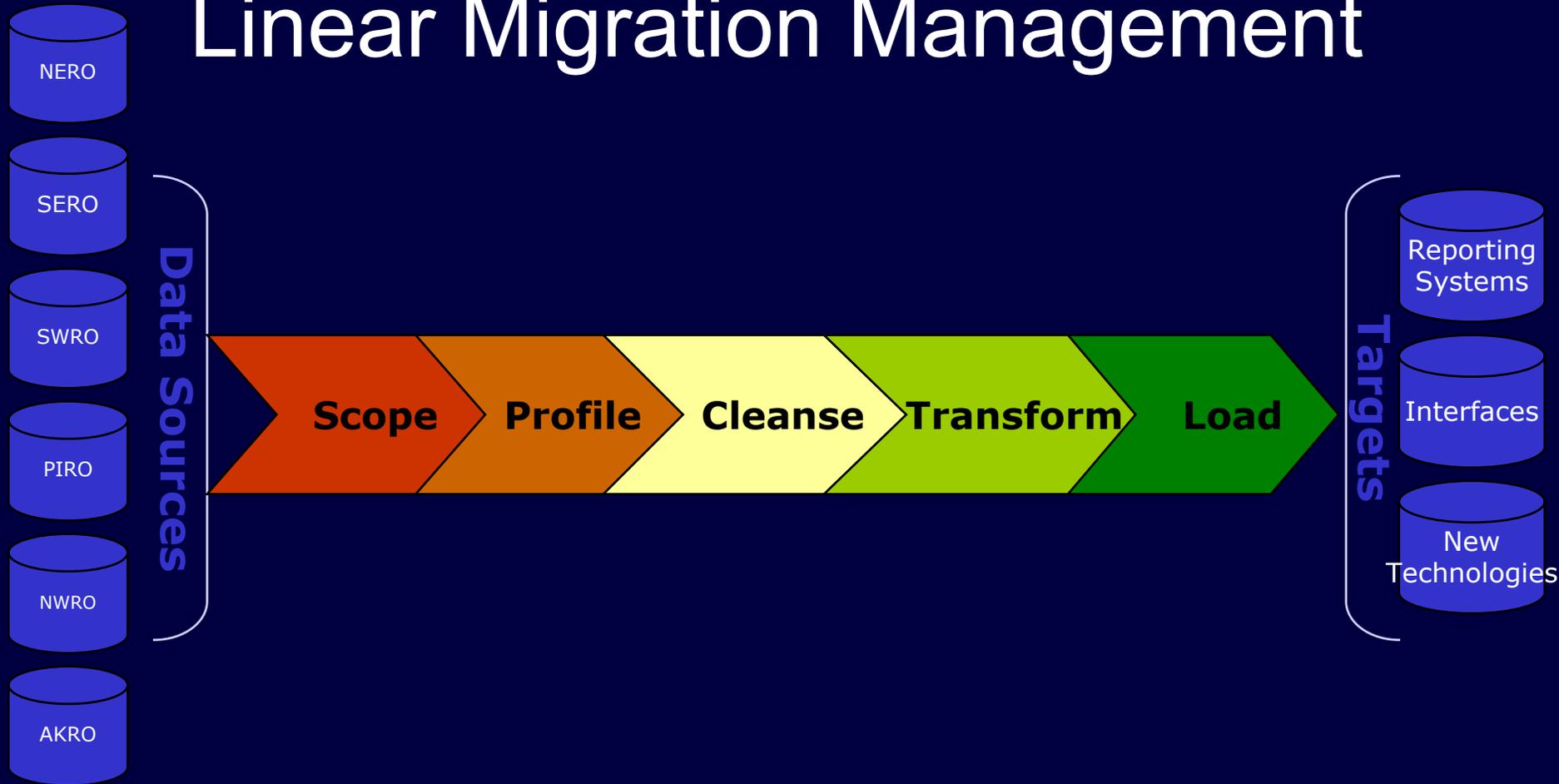


# The Solution





# Linear Migration Management





# Linear vs. the Traditional Approach

Linear Approach
  Traditional Approach





# Keys to Data Analysis

## Profile Statistics

POP, Null, Blank,  
Min, Max

## Pattern Analysis

Format Trending

## Value Distribution

Value Trending





# Keys to a Successful Migration

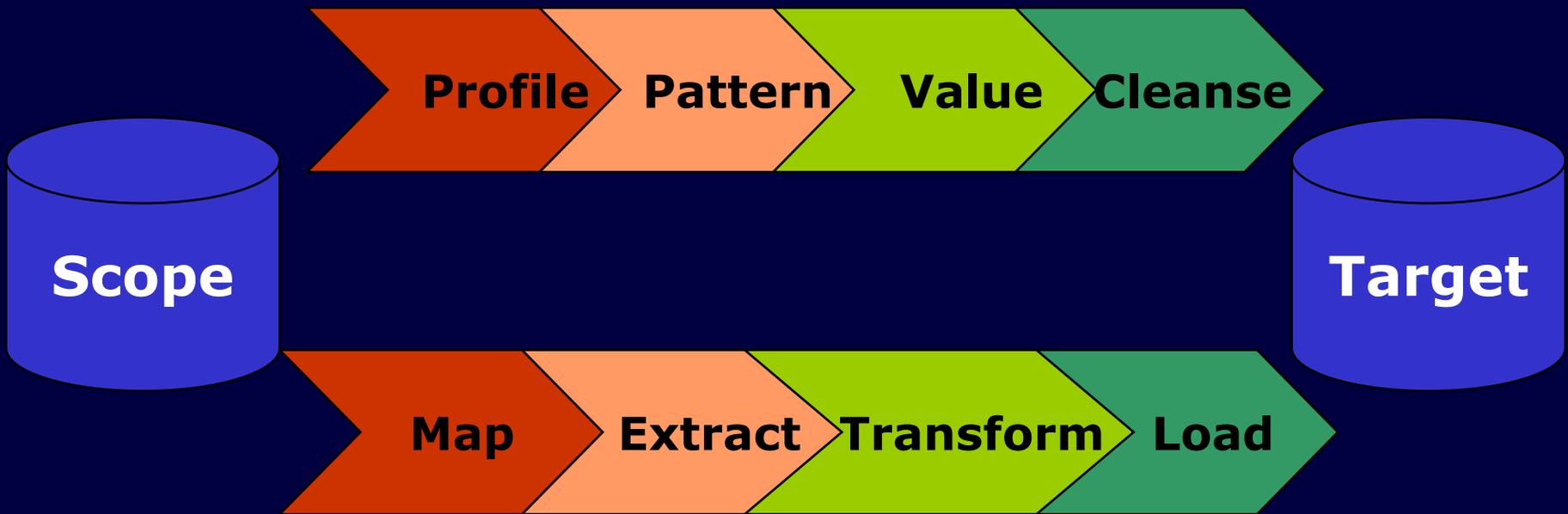
- Separate the good data from the bad
- Separate cleansing from transforming





# Linear Migration Management

## Structured Analysis, Improved Workflow





# Summary

- Poor data quality costs more than \$600 billion annually (TDWI)
- To avoid “code, load and explode,” create a migration process that integrates:
  - Data profiling – Analyze data and uncover potential problems
  - Data quality – Cleanse, validate and correct data errors





# Questions and Answers

Joe Hudicka





# SWR Harpoon Fishery Data Quality Analysis

**Eric Pilkington**

